International Journal of Computer Science and Engineering Communications- IJCSEC. Vol.2 Issue.2, April 2014. ISSN: 2347–8586

Detection of Breast Mass in Digital Mammogram from Variable Hidden Neuron Ensemble Based Technique of Mass Classification Using Region Growing Segmentation

Geetha .K.R, Nanthini.K, Jeevitha.S, Abirami.G
Department of Computer Science and Engineering,
Professional Group of Institutions, Palladam, India
krgeethu4393@gmail.com, nandnigreen@gmail.com

Abstract: Digital mammograms are the best method to detect the breast cancer in earlier stage using image processing methods. In this paper a new technology that enhances the variable hidden neural network for detecting the location of breast mass is proposed. First, the pre-processing methods are performed over digital mammogram image. Then the ROI is extracted from the pre-processed image. The Region Growing Segmentation is implemented to separate the part of the image that having the same pixel values from the mammogram. After that the features such as density, mass shape, mass margin, Abnormality Assessment rank, patient age, Subtlety value are extracted. The next process starts off with the creation of the neural networks by varying the number of neurons in the hidden layer. These are then trained, tested and ranked according to the classification of accuracy. To create an ensemble network, the Ten-Fold Cross validation which produces the classifiers, is used. The classifiers are then fused together to create the final ensemble network which reveals whether the image is malignant or benign.

Keywords: *Breast cancer, Mammogram, Ensemble neural network, region growing segmentation, adaptive histogram equalization.*

INTRODUCTION: Breast cancer affects large number of women population. The breast cancer is mainly occurred in the inner lining of the lobules which supplies the milk and the milk ducts which carries the milk from the lobules to the nipples and they called as lobular carcinoma and ductal carcinoma respectively. There are many factors that cause breast cancer that are calcium deposition in the breast, radiation exposure, obesity, getting aged, genetic problems and consumption of alcohol. According to the survey taken by the National Cancer Institute, 232240 females and 2240 males affected by breast cancer yearly in USA. Among them 39620 were died. In Australia, one in nine women is diagnosed with the breast cancer in their lifetime [1]. The various methods such as examining the breast, breast ultrasound, breast MRI, biopsy, mammograms, 2D combined with 3D mammograms are used to detect the breast cancer. The mammogram is taking an X-ray by compressing the breast between the two plastic plates [2]. The mammogram gives the better visibility at the skin, greater image flexibility, shorter exam times and more confidence in the results. The image processing is a physical process that takes an image as the input and produces an image or the parameters related to the image as the output. Many computer vision and computational intelligence based techniques are developed in past 20 years. It has a main disadvantage that a consistent and acceptable accuracy has not been achieved. Then Artificial Neural Networks (ANN) has been successful and demonstrated better than the other traditional methods [3]. The ANN is proposed as a simulation of the central nervous system of the human. The artificial neural networks are non-linear information processing device that are built from the interconnected neurons. The modern computers use an algorithmic approach to solve a specific problem but ANN process the information in a similar way that the human brain does. But it has low classification accuracy [4]. Recently, the ensemble techniques have been applied and shown that they have achieved higher accuracy than a single neural network. The ensemble technique is mainly based on the diverse base classifiers which produces better result. The performance of the ensemble technique is improved by the diversity. The diversity is introduced by varying the number of neurons in the hidden layer of the neural network [5]. The ensemble technique distinguishes the similar characteristics of benign and malignant breast masses.

Paper Type: Research
Corresponding Author: Geetha .K.R, email: krgeethu4393@gmail.com
Online: www.scientistlink.com

RELATED WORKS

Gou et al [6] proposed technique called Partition Based Network which is a boosting technique for class imbalanced datasets. They train the ensemble to classes that convert the network training into balanced training problem. This technique achieved a classification accuracy of 96.50%. Liu et al [7] proposed a technique called random forest which is an ensemble technique that achieved a classification accuracy of 79%. Roselin et al [8] proposed a meta-heuristic ensemble classifier technique which uses the ant miner and ten-fold cross validation on MIAS dataset. This method achieved the classification accuracy of 83%. Meena et al [9] proposed a modular neural network technique that evaluates on the Wisconsin breast cancer dataset from UCI online repository. They achieved a classification accuracy of 96.87%. Huang et al [10] proposed clustering ensembles based on multi classifiers fusion. They achieved a classification accuracy of 89.50%. Kinnard et al [11] utilized the region growing with the analysis for segmentation contours and a multiple circular path convolution neural network for determining the cancer as malignant or benign. Costa et al [12] proposed a new coding technique that utilizes the tenfold cross validation and they achieved a classification accuracy of 90.07%. Other methods to obtain the diverse classifiers are the Bootstrapping and Adaboost methods. Melville and Mooney created a DECORATE mechanism to generate the diverse ensembles. Li et al [14] proposed a feature weighting framework to produce the diverse classifiers. Peter Mc Leod and Brijesh Verma proposed the Variable hidden neuron ensemble technique of classification and achieved classification accuracy of 98%. This method distinguishes between malignant and benign masses. It does not use any segmentation method to detect the location of the breast mass.

OVERVIEW

This paper proposes a technique called Region Growing Segmentation for segmenting the breast mass from the digital mammogram and the variable hidden neuron ensemble technique for the mass classification. The pre-processing methods and the Region Growing Segmentation are applied to the digital mammogram. The Region of Interest and the features are extracted from the pre-processed image. The ensemble network is created by training the neural network with varying the number of neurons in the hidden layer, choosing the best performers, fusing all the results together and determining whether it is malignant or benign. The modules in the proposed technique are explained below.

PREPROCESSING METHODS

Acquiring of digital mammograms

The digital mammograms were acquired from the database called Digital Database of Screening Mammography [DDSM]. The mammograms are available in the website: http://marathon.csee.usf.edu/Mammography/Database.html. This database contains 2620 cases in 43 volumes.

Image Reading

The MATLAB environment represents the binary image as one dimensional array, grayscale image as two dimensional array and the color image [RGB] as three dimensional array (one 2-dimensional array for each of the color plane). The size of the image is represented by two factors such as height (number of rows of the array) and width (number of columns of the array). By pointing the z-axis to the front of the image, the x and y co-ordinates are chosen. A pixel is the single point in the image [15].

Image scaling

Image scaling is the important process in image processing and image analysis [16]. Image scaling is the non-trivial process of resizing the digital image that involves smoothness and sharpness. The various scaling methods are nearest neighbor interpolation, bilinear interpolation and super sampling.

Adaptive histogram equalization (AHE):

It is a technique in computer image processing to improve contrast in images. It computes many histograms; each corresponds to a different part of the image, and then uses them to redistribute the lightness values of the image. The ordinary histogram equalization process uses only a single histogram for an entire image [15]. Adaptive histogram equalization is considered as an image enhancement technique to improve an image's local contrast, and to bring out more details in the image.

BLOCK DIAGRAM

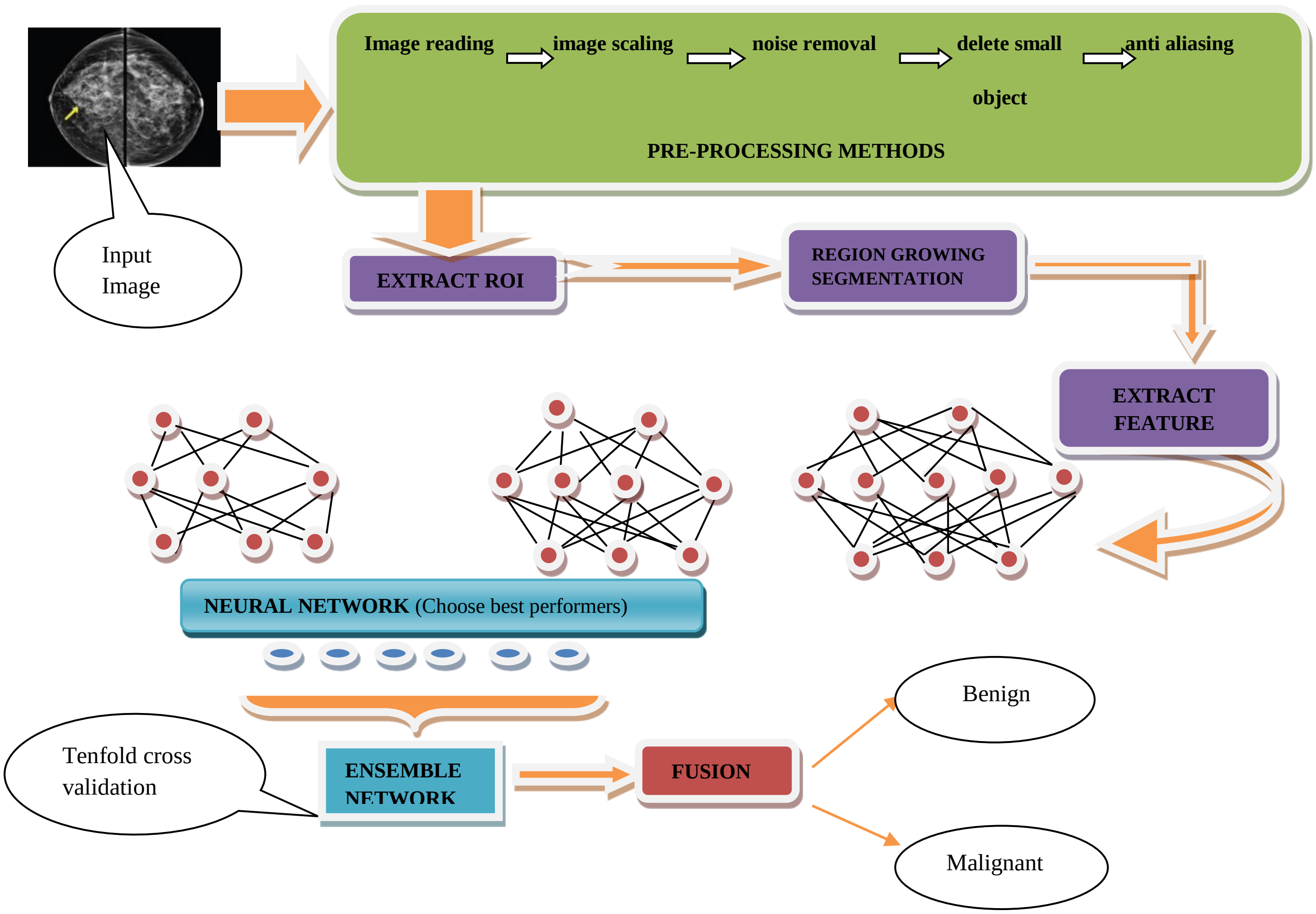

Figure 1. The ensemble network with region growing segmentation

Removing noise: The noise in the image is normally due to the environment conditions, sensor quality and human interference.

Deleting the small objects: It works with one main object and filter out all the remaining objects. This is done through deleting any object that has a size below the size of the largest object in the image (after sorting). Deleting an object essentially makes all of its pixels in the image matrix given a value of false.

Region of Interest: The ROI are extracted based on information provided in the form of a chain code by radiologists. The DDSM [8] provides a chain code that allows for the segmentation and extraction of ROIs. This shows the process right from the initial mammogram through to final classification. The chain code allows for the anomaly to be extracted by locating the start coordinate and then working through the chain code in order to extract the anomaly and the surrounding boundary tissue. The chain code extraction starting point has been identified, and the boundary would be walked in order to complete the extraction.

Feature Extraction

An anomaly cannot be mapped to diagnosis without utilizing the certain features. This research utilizes the same type of features that could be used as training tool for conceptual understanding.

Some of features are as follows [17,19]

- Density
- Mass shape
- Abnormality Assessment rank
- Patient age
- Subtlety value

Density The BI-RADS reporting system is used to measure the density of a mass anomaly. If masses of a density is equivalent to surrounding tissue are harder to identification

Mass shape: For a diagnosis the shape of a mass is very important. Benign masses have distinct margins and more compact in nature. The malignant shapes are irregular with hard to define margins.

Abnormality Assessment rank This is an assessment of how serious the anomaly is on a one to five category rating where one indicates that it is not likely to be malignant and five is highly suggestive of malignancy.

Patient age: The breast cancer frequency is increased with age. Some researchers note that more aggressive cancers occur in younger women where the cancer is harder to detect and diagnose.

Subtlety value: This represents how difficult to find the lesion.

REGION GROWING SEGMENTATION

Region is a group of connected pixels with the similar properties. The image is partitioned into regions by the use of gray values of the image pixels. The two general approaches for portioning the images are Region-based segmentation Boundary estimation using edge detection [18].

The basic formulation is, given a set of image pixels I and a homogeneity predicate P(.), find a partition S of the image I into a set of *n* regions R_i such that

$$\bigcup_{i=1}^{n} Ri = True$$

$$P(Ri) = True \quad , \ for\ all\ i$$

i.e any region satisfies the homogeneity predicate

Any two adjacent regions cannot be merged into a single region

$$P\left(Ri \bigcup Rj\right) = False$$

The main goal of segmentation is to partition an image into regions. Some segmentation methods such as thresholding achieve this goal by looking for the boundaries between regions based on discontinuities in gray levels or color properties. Region-based segmentation is a technique for determining the region directly. Adaptive Thresholding:- Used in scenes with uneven illumination where same threshold value not usable throughout complete image. In such case, look at small regions in the image and obtain thresholds for individual sub-images. Final segmentation is the union of the regions of sub-images. Variable Thresholding: Approximates the intensity values by a simple function such as a plane or biquadrate. It is called background normalization.

VARIABLE HIDDEN NEURON BASED ENSEMBLE TECHNIQUE

The ensemble technique is created by varying the internal architecture of neural network and input data. This concept produces diverse neural network classifiers and it is combined using hierarchical fusion. The neural network obtains the knowledge about the problem domain by training it on the population sample. The diversity will be created only if the knowledge about internal architecture and parameters of each neural network classifier are represented in different way by each classifier. The learning of different characteristics of masses in digital mammogram is allowed by the classifier associated with this diversity. Many techniques have been used by the researchers for manipulating the training dataset to introduce diversity into the ensemble. The investigation has not been fully completed to use different types of neural network with varied hidden neurons and input data in ensemble creation. The traversal of feature space is caused by varying the number of neurons which is different to result in different weight values. Because of this each neural network obtains different behavior. It results in introducing diversity into the resultant ensemble. Here the effect obtained on the creation of ensemble is found by varying the number of hidden neurons in the hidden layer of neural network at the same time tenfold cross validation is used. Two scientists named Partridge and Yates introduced diversity by using small number of neurons in the research. Here it is different because wide range of neurons and tenfold cross validation is used it introduces diversity into the ensemble. Another method hierarchical fusion is used and ensemble has been developed for mass classification.

GENERATION OF ENSEMBLE NETWORK

The ensemble network is generated for the classification of masses in which different classifier with different neural network architecture and input data learn different characteristics. The neural network is created by varying the number of neurons in the hidden layer. The training process is done following this testing and ranking process is done according to classification accuracy. The performance has been improved in terms of accuracy and consistency by combining the different decision produced for the same input by the networks. The digital mammogram has different characteristics in some areas so because of this the ensemble can learn and generalize better than individual classifier. The ensemble network is created using neural network in which the numbers of hidden neurons in the neural networks were incremented this result in generation of constituent classifier. The majority vote algorithm is used to create the final ensemble network by combining the classifiers together. Here the hidden neurons were varied from 2 to N. The value of N is 150 in this study and the further investigation has to be carried out to find the suitable value of maximum neurons. The neural networks are trained using ten-fold cross validation. After completing the training process the best performing neural networks are selected. The final ensemble network is created by using the selected neural networks.

EXPERIMENTAL RESULTS AND ANALYSIS

Many experiments have been conducted by using the dataset obtained from the DDSM benchmark database [17]. DDSM database contains 200 mass anomalies which are classified between malignant and benign cases. The ensemble technique and region growing segmentation has been implemented. For each individual classifier the experiments were conducted. The experiments were also conducted for ensemble classifier without ten-fold cross validation combined with another ensemble classifier called ADABOOST M1 [13] for comparison purpose. The individual classifiers have shown the classification accuracy of 83%for lowest value and 86% for highest value in the single layer back propagation neural network. The lowest and highest classification accuracy shown by ensemble technique was 93% and 98%. To detect the effect of ten-fold cross validation on the creation of ensemble the experimental results can also be obtained using the same dataset without ten-fold cross validation this involves the splitting of dataset into 50% training data and 50% test data. ADABOOST is one of the iterative mechanism in which each classification problem is associated with weighting. The classification problem having higher weighting will be harder to classify. Through the weighted vote of the constituent classifiers the final classification is constructed. The region growing segmentation is used along with the ensemble network to find the cancer affected area accurately. Comparing to other techniques like MLP, ensemble MLP, and ADABOOST M1 networks the proposed ensemble technique has shown higher accuracy of 98%. By reducing the variance in prediction errors the ensemble shows higher accuracy. The classification accuracy has been improved by varying the number of neurons in the hidden layer simultaneously ten-fold cross validation is used. The ANOVA analysis was conducted to prove the improvement of classification accuracy obtained between the neural network and ensemble technique.

CONCLUSION AND FUTURE WORK

In this paper new approach have been used for analyzing and classifying masses in digital mammograms. The ensemble technique uses variable hidden neurons hierarchical fusion and ten-fold cross validation and it is evaluated on a subset of DDSM benchmark database. The experimental result has shown classification accuracy of 98% over single neural networks where ADABOOST has obtained classification accuracy of 86% and 90%. Many comparative analyses have been conducted in which ensemble technique provides better results than the existing techniques for the purpose of classification of masses in digital mammograms. The region growing segmentation is used to find the accurate location affected by cancer.

The research shown in this paper has to be examined to understand the behavior of variable hidden neuron based ensemble technique for mass classification. The classification performance without the weighted diversity measures is used. For larger ensembles it explains the classification accuracy. The further investigation has to be done on the impact of number of neurons and number of individual network on accuracy. The future research will be done to find the highest suitable number.

ACKNOWLEDGMENT

We would thank our management and faculty members for this kind initiation of our project.

REFERENCES

[1] South Jersey Radiology Associates. (2012, Sept. 29). *Women's Imaging. PA*. [Online]. Available: http://v002-dt. www.comm.com/imaging/serv-digital.asp

[2]Robin J and Mason P (2002) Handbook of pharmacy healthcare (diseases and patient advice). *Pharmaceutical Press*, London, UK. ISBN 0- 85369-507-5.

[3] U. Rani, "Parallel approach for diagnosis of breast cancer using neural network technique", Int. J. Comput. Applicat, vol. 10, no. 3, pp. 1–5, Nov. 2010.

[4] M. Abdelaal, H. Sena, M. Farouq, and A. Salem, "Using pattern recognition approach for providing second opinion of breast cancer diagnosis," in Proc. 2010 7th Int. Conf. Informatics Systems, pp. 1–7.

[5] E. Santos, R. Sbourin, and P. Maupin, "Pareto analysis for the selection of classifier ensembles," in Proc. 10th Annu. Conf. Genetic Evolutionary Computation, 2008, pp. 681–688.

[6] S. Gou, H Yang, L. Jiao, and X. Zhuang, "Algorithm of partition based network boosting for imbalanced data classification," in Proc. 2010 Int. Joint Conf. Neural Networks, pp. 1–6.

[7] J. Liu, J. Chen, X. Liu, and J Tang, "An investigation of mass diagnosis in mammogram with random forest," in IEEE Proc. 4th Int. Workshop Advanced Computational Intelligence, 2011, pp. 638–641.

[8] R. Roselin and K. Thangavel, "Classification ensemble for mammograms using ant-miner," in Proc. IEEE 2nd Int. conf. Computing, Communication Networking Technologies, 2010, pp. 1–6.

[9] Y. Meena, K. Arya, and R. Kala, "Classification using redundant mapping in modular neural networks," in Proc. IEEE 2nd World Congr. Nature Biologically Inspired Computing, 2010, pp. 554–559.

[10] Y. Huang, D. Monekosso, and H. Wang "Clustering ensemble based on multi-classifier fusion," in Proc. 2010 IEEE Int. Conf. Intelligent Computing Intelligent Systems, pp. 393–397.

[11] L.M. Kinnard, S.C.B. Lo, P.C. Wang, M.T. Freedman, M.F. Chouikha, "Separation of malignant and benign masses uses maximum-likelihood modeling and neural networks," in Medical Imaging 2002: Image Processing, San Diego, CA, 2002, pp. 733-741.

[12] D. Costa, L. Campos, and A. Barros, "Classification of breast tissue in mammograms using efficient coding," BioMed. Eng. Online, vol. 10, no. 55, 2011.

[13] M. Skurichina and R. Duin, "Bagging, boosting and the random subspace method for linear classifiers," Pattern Anal. Applicat., vol. 5, no. 2, pp. 121–135, 2002.

[14] Y. Li, S. Gao, and S. Chen, "Ensemble feature weighting based on local learning and diversity," in Proc. 26th AAAI Conf. Artificial Intelligence, 2012, pp. 1019–1025.

[15] Thaler M and Hochreutener H (2008) Image processing basics using MATLAB.

[16] Gonzalez R, Woods R and Eddins S (2004) Digital images processing. 2nd edition, Prentice-Hall, Inc. NJ, 07458, ISBN 0-13-008519-7.

[17] R. Panchal, "Significance of features in classification of breast abnormality patterns in digital mammography," Ph.D. Dissertation, Central Queensland Univ., Brisbane, Australia, 2006.

[18] "A Parallel Computer System for the Detection and Classification of Breast Masses" Soha Ekrima K. Yousuf , Sahar Haj Ali A. Mohammed Department of Engineering, Faculty of Biomedical Engineering University of Medical Sciences and Technology.

[19] S. Luo and B. Cheng, "Diagnosing breast masses in digital mammography using feature selection and ensemble methods," J. Med. Syst., vol. 36, no. 2, pp. 569–577, Apr. 2012.

[20] T. Andre and R. Rangayyan, "Classification of breast masses in mammograms using neural networks with shape, edge sharpness, and texture features," J. Electron. Imag, vol. 15, no. 1, p. 013019, 2005.